

L'écrit et le document

Segmentation d'images de documents par une approche multirésolution. Extraction précise des lignes de texte

Document Image Segmentation Using a Multiresolution Approach. Accurate Text Line Extraction

par O. DEFORGES, P. PIQUIN, C. VIARD-GAUDIN, D. BARBA

Laboratoire SEI — EP CNRS 0063/IRESTE
Route de Gachet
F-44087 Nantes cedex 03

Résumé

Cet article présente une méthodologie et des outils de traitement permettant de localiser puis d'extraire précisément les lignes de texte contenues dans l'image d'un document. La classe des documents visés est de type document très complexe, leurs contenus étant totalement non contraints. Globalement la méthodologie s'articule autour de trois étapes clés. La première est une localisation des zones d'intérêt. Elle est réalisée directement sur l'image en niveaux de gris et utilise une approche multirésolution garantissant une grande robustesse vis-à-vis de la très forte variabilité des textes : taille, disposition, présentation. Une étape de binarisation réalisée séparément pour chaque zone d'intérêt permet dans une seconde phase l'extraction proprement dite des lignes de texte. Enfin, une post-segmentation faisant coopérer la localisation initiale et des caractéristiques structurelles extraites de la ligne permet de traiter les cas très perturbants pour la lecture du chevauchement de lignes sur de l'écriture manuscrite. Des exemples relevant de la problématique de l'extraction des lignes du bloc adresse sur objets postaux (grandes lettres, magazines) illustrent cet article.

Mots clés : Images de documents, segmentation, multirésolution, extraction de lignes de texte.

Abstract

An overall scheme and related algorithms performing accurate text lines extraction from an image of document are described in this paper. The type of documents concerned here is very complex, with totally unconstrained data. Postal objects, especially the so-called flat objects, i.e. : large envelopes, magazines,... are within this kind of documents. Three main phases have been considered to achieve the overall function. First of all, areas of interest are located using a multiresolution approach allowing to preserve from large variability of text features. This is performed directly on the gray-level image. A binarization stage, taking advantage of the results of the localization, is next performed to extract the lines. At last, a post-segmentation involving the located areas in the gray-level images and structural features extracted from the lines allows to deal with severe cases such as overlapping lines induced by handwritten texts. Examples related to text line extraction on postal objects are illustrating this paper.

Key words : Document image analysis, segmentation, multiresolution, text line extraction.

1. Introduction

L'analyse automatique de documents est un champ vaste d'applications potentielles compte tenu de la quantité énorme de documents produits par l'activité humaine. L'utilisation effective de système automatique d'analyse de documents papiers dépend évidemment des capacités à localiser les zones informatives et à les reconnaître. Cependant, les résultats obtenus aujourd'hui en reconnaissance optique de caractères ne sont satisfaisants qu'au

prix de conditions très restrictives si l'on considère la très grande variété des documents existants. En particulier, le texte doit être en caractères imprimés, plusieurs polices étant possibles. Il doit être de bonne qualité et disposé régulièrement — typiquement en disposition Manhattan — sur un fond uniforme. Or beaucoup d'applications potentielles sortent de ces limites : orientation variables des lignes, fond incontrôlé, écriture manuscrite, mélange de textes, de graphiques, de photographies. C'est le cas des documents techniques, des cartes et plans, des magazines où la mise en page peut être très complexe, ou encore des documents postaux

où l'adresse peut n'être qu'une faible fraction de l'information présente.

Dès lors une étape essentielle d'un système complet d'analyse, en amont de la lecture elle-même, consiste à localiser et extraire le plus correctement possible toutes les entités à traiter par le système de reconnaissance. C'est dans ce cadre que se situent les travaux présentés dans cet article.

Les approches utilisées jusqu'ici pour traiter ce problème consistaient à binariser dès l'acquisition l'image du document en utilisant un algorithme plus ou moins évolué de binarisation. Suivant le type de document et la méthode employée [1], le résultat n'est pas toujours bon. A partir de l'image binaire obtenue, des méthodes de segmentation sont alors mises en œuvre. On trouvera dans [2] un exposé de synthèse sur de telles méthodes. Rappelons brièvement ici que deux types d'approches sont classiquement rencontrées. La première est intéressante lorsque des connaissances sur la structure du document sont disponibles, le modèle sous-jacent guide alors la segmentation. C'est le cas des méthodes dites descendantes dont l'algorithme RXYC (Recursive XY Cut) basé sur des profils de projection est l'une des références en la matière [3]. La deuxième catégorie d'approches, de type ascendant, est plutôt guidée par les données : elle procède par regroupements progressifs d'entités élémentaires partant du pixel pour construire les structures physiques du document. Une des grosses difficultés est de savoir alors définir les règles permettant de contrôler les fusions. Il a été proposé récemment [4] pour cela un schéma incluant une étape de relaxation pour valider la formation des lignes et des blocs de lignes. Ce type de traitement est malheureusement très coûteux en temps machine : typiquement plusieurs minutes actuellement sur station de travail. De plus, la plupart des méthodes évoquées ci-dessus sont destinées à traiter des documents imprimés, rares sont les contributions dédiées à l'écriture manuscrite [5].

2. Schéma général

Le schéma traditionnel : binarisation, segmentation puis lecture, a du être reconsidéré de façon à pouvoir tenir compte de la grande variabilité des contextes d'extraction.

La structure générale du système de traitement proposé est présentée figure 1. La méthode de localisation opère directement sur l'image en niveaux de gris. Elle permet de détecter toutes les zones de texte quels que soient la taille des caractères, l'orientation des lignes et le style de l'écriture. De manière complémentaire, les éléments de type trait sont également reconnus. Une approche multi-résolution se montre particulièrement appropriée pour traiter ce problème, celle-ci est présentée dans la section 3. On présente ensuite, dans la section 4, des exemples illustrant la méthode dans le cas de l'application à la localisation du bloc adresse d'objet postaux. La binarisation est ensuite effectuée localement de façon adaptée sur chaque zone d'intérêt. L'intérêt

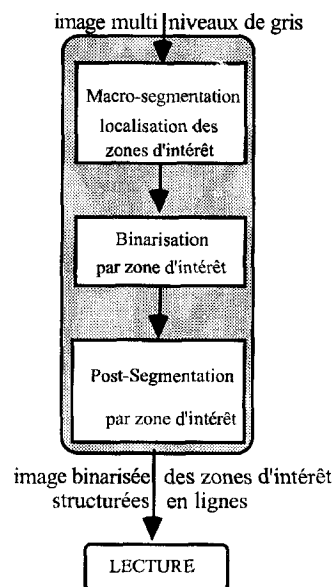


Figure 1. – Schéma global de la méthode de traitement.

d'une telle démarche est présenté dans la section 5 à l'aide d'exemples. Enfin, on définit la phase de post-segmentation dans la section 6. Elle permet de raffiner les résultats de la localisation et de résoudre les problèmes de chevauchement de lignes qui sont fréquents dans le cas de l'écriture manuscrite et très perturbants pour la reconnaissance.

Une base de plus de 200 images de tests provenant du S.R.T.P. (Service de Recherche Technique de la Poste) a permis de valider l'ensemble de l'approche.

3. Localisation par approche multirésolution

L'originalité de la méthode de localisation tient dans la modélisation même du texte, servant à son extraction. Le principe est ici de mettre en évidence la structure linéaire globale du texte, perceptible à basse résolution. L'idée sous-jacente est qu'en observant une ligne de texte de « suffisamment loin », la perception que l'on en a est proche de celle d'un trait plein. Cette façon de procéder est à rapprocher de l'une des rares méthodes faisant référence au fonctionnement du système visuel humain et opérant sur des images multi-niveaux de gris [6] où des caractères de grandes dimensions sont localisés à l'aide d'une mise en évidence des traits verticaux aux frontières des caractères, un filtre d'analyse de largeur variable induisant l'aspect multirésolution. Le recours à l'aspect basse résolution de l'image sur lequel nous nous appuyons permet d'effectuer les traitements sur une image à la résolution très réduite de 75 dpi (dots per inch) comparativement à celle nécessaire à la lecture qui est typiquement de 200 dpi.

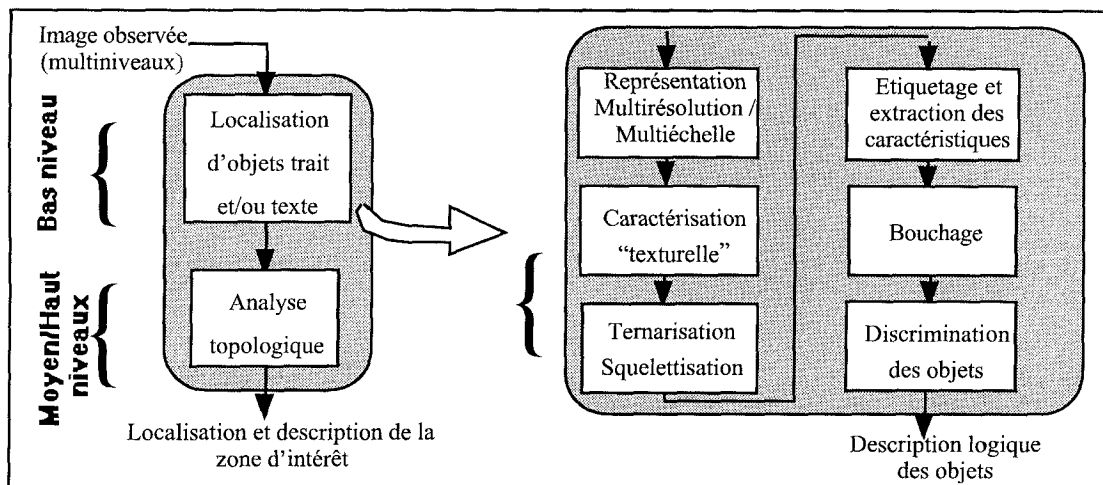


Figure 2. – Blocs diagrammes du système complet et des traitements de bas niveau.

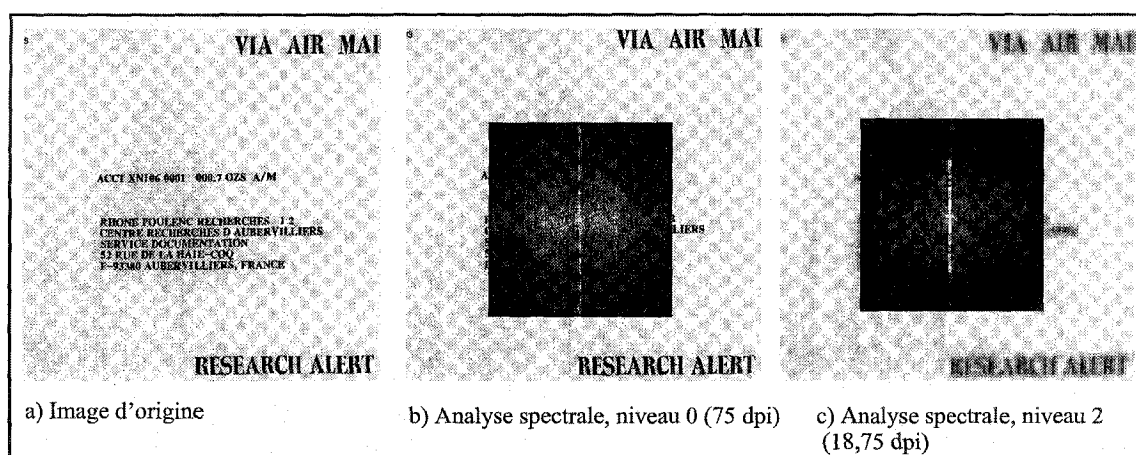


Figure 3. – Caractérisation fréquentielle du texte à différentes résolutions.

Le schéma bloc général du système de localisation de zones d'intérêt est défini figure 2. Seule la partie des traitements de bas niveau permettant d'extraire de manière complémentaire les objets de type texte et de type trait va être décrite dans la suite. Elle représente en effet l'aspect le plus novateur en tant que méthode de segmentation sans étape de binarisation initiale. Elle s'avère également très performante d'un point de vue rapport efficacité/complexité opératoire par le caractère local et la régularité des traitements mis en œuvre.

3.1. CARACTÉRISATION TEXTUELLE MULTIRÉSOLUTION

L'utilisation d'une représentation pyramidale de type pyramide Gaussienne [7] va présenter deux propriétés essentielles concernant les champs d'applications envisagés :

- aptitude dans une large mesure à s'affranchir de la grande dynamique de la hauteur des caractères (aspect multiéchelle) [8].
- mise en évidence de structures globales des entités (aspect multirésolution) [9].

La pyramide Gaussienne est construite itérativement à partir d'un filtre Gaussien 1-D appliqué suivant les directions horizontale et verticale, et suivi d'une décimation dyadique (un pixel sur deux retenu, suivant les deux directions principales). Cette structure pyramidale induit un coût d'occupation mémoire supplémentaire de 33% au maximum. La figure 3 présente une image originale et le résultat d'une transformée de Fourier effectuée dans une fenêtre de l'image à différentes résolutions. Il apparaît clairement que lorsque la résolution est suffisamment faible, l'aspect linéique du texte est prédominant (cela se traduit par de fortes valeurs dans le plan fréquentiel le long de l'axe vertical).

L'hypothèse posée, et qui se trouve vérifiée par l'expérience, est que l'utilisation d'un facteur de décimation dyadique entre les niveaux de la pyramide est suffisant pour garantir au moins que dans un niveau de la pyramide, pour peu que le nombre de niveaux soit suffisant, le texte apparaisse sous une forme linéique.

Une conclusion évidente d'une telle observation est qu'il est possible de construire un schéma d'analyse permettant l'extraction des lignes de texte et celle des traits. Percus de la même façon à un niveau adéquat, les deux types d'entités porteront alors le dénominateur commun d'*objets linéiques*. Ils partagent ce que nous appelons un premier attribut distinctif. Parmi les autres types d'informations potentiellement présentes dans les images, les plus fréquentes sont les graphismes, des textures plus ou moins complexes, et du bruit de nature aléatoire ou structurée (par exemple le reflet des enveloppes plastiques). La mise en évidence dès la première phase de l'analyse d'un attribut spécifique à la classe d'objets recherchés permet d'être beaucoup plus robuste au bruit en comparaison avec des méthodes où l'opération de binarisation s'effectue de manière préalable.

Si le problème d'extraction des objets linéiques est posé en termes de caractérisation de texture [10], le but peut être alors de déterminer la transformation linéaire locale la mieux adaptée à la texture à analyser. N'ayant aucun a priori sur l'orientation des documents et donc des informations écrites qui s'y trouvent, deux directions principales ont été définies pour les traitements. Ainsi, deux pyramides de caractérisations orthogonales sont obtenues à l'issue de ces phases de filtrage : une appelée *pyramide à caractérisation horizontale*, l'autre *pyramide à caractérisation verticale*.

L'a priori sur la forme linéaire des objets à détecter a conduit à utiliser des jeux de filtres séparables 2-D proches de ceux adoptés dans [11] : un filtre passe-bas 1-D dans la direction supposée des lignes permettant de mieux fusionner les caractères et les mots, et un filtre 1-D de type Laplacien dans la direction orthogonale afin de détecter les bords (cf. figure 4).

Une telle approche par filtrage à large bande passante de fréquences spatiales présente l'atout majeur d'être beaucoup plus robuste à l'inclinaison que les méthodes classiques de détection du texte (RXYC, RLSA).

La définition de deux pyramides de caractérisation se trouve justifiée non seulement par le manque d'information quant à l'orientation, mais aussi parce que ces deux pyramides vont se révéler complémentaires ultérieurement pour différencier par un second attribut les éléments texte de ceux des traits.

Dans la suite des traitements de bas niveaux, les traitements sont effectués indépendamment à chaque niveau de chacune des deux pyramides.

Les traitements de niveaux supérieurs, et suivant les deux directions, ont ensuite à fusionner cette description multirésolution.

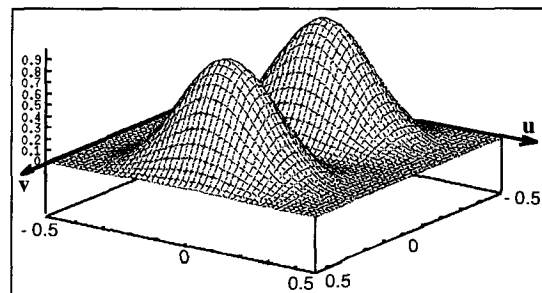


Figure 4. – Réponse fréquentielle du filtre d'analyse 2-D (caractérisation horizontale).

3.2. TERNARISATION — SQUELETTISATION

La première phase de l'extraction proprement dite consiste à ne retenir que l'information utile. Celle-ci correspond ici à la localisation des bords supérieur et inférieur des objets. La distinction de la position de ces bords permettra par la suite une reconstruction des lignes médianes du texte sans ambiguïté. Ainsi, à chaque pixel va être affectée une des trois valeurs possibles, réalisant alors une opération de quantification à trois niveaux (ternarisation) :

- 0 → *fond*
- 1 → Transition négative, correspondant au *bord supérieur* des objets.
- +1 → Transition positive, correspondant au *bord inférieur* des objets.

Dans notre problématique où l'on ne s'attache à trouver que des éléments de forme linéique, il est particulièrement avantageux d'obtenir des éléments d'épaisseur unitaire. Par simplicité de notation, nous nous plaçons ici dans l'hypothèse de la pyramide à caractérisation horizontale. La méthode est analogue pour la caractérisation verticale à une rotation près.

Les deux étapes de ternarisation et de squelettisation sont ici couplées. La ternarisation va s'appuyer sur la pyramide de caractérisation (horizontale et verticale indépendamment) en y détectant les passages par zéro significatifs dans la direction du filtrage Laplacien. Les pixels correspondants sont alors marqués selon le signe de leur gradient calculé suivant cette même direction.

La squelettisation consiste à assurer localement l'épaisseur unitaire des composantes. Les pentes locales étant bornées entre $[-45^\circ; +45^\circ]$ du fait des propriétés des filtres utilisés, les relations d'adjacence entre deux pixels marqués Transition de même type sont interprétées selon la figure 5. Une configuration telle que celle de la figure 5-d) ne constitue pas une information significative, mais plutôt une redondance entre les deux pixels (épaississement du trait). Aussi, ne sera conservé comme seul point de transition que le pixel présentant la valeur la plus proche de zéro

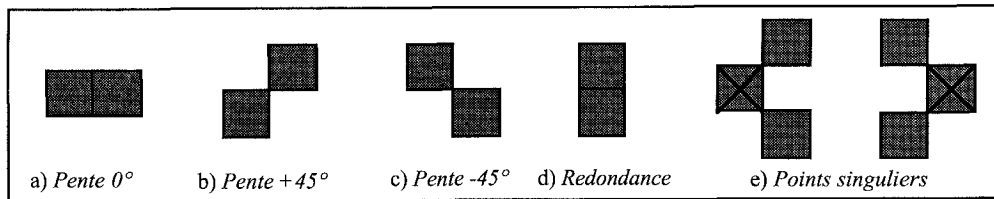


Figure 5. – Interprétation des pentes locales.

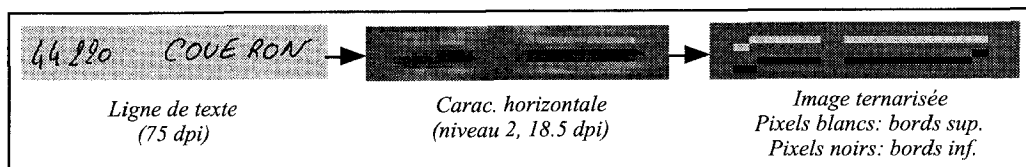


Figure 6. – Extraction des bords d'une ligne de texte.

dans l'image filtrée. Les composantes sur leur totalité vont donc effectivement être squelettisées dans la direction perpendiculaire à la direction de caractérisation.

Cette opération combinée Ternarisation–Squelettisation est réalisée conjointement en un seul balayage de l'image, de façon indépendante colonne par colonne.

Du fait des limitations de pente, il est impossible de trouver des valeurs absolues de pentes supérieures à 45° , sauf si des configurations telles que sur la figure 5–e) se présentaient. Les points marqués d'une croix sont appelés *points singuliers*. Ils correspondent à des « points de retour » ou « dédoublement » des composantes. En fait, le principe de notre ternarisation empêche la création de tels points, du fait de l'alternance des signes des pixels.

Une étape complémentaire de bouchage courte-distance de type morphologique est ensuite mise en œuvre. Un exemple de résultats du procédé d'extraction des bords est donné figure 6. Les pixels bord supérieur apparaissent en blanc, bord inférieur en noir.

3.3. ÉTIQUETAGE ET EXTRACTION DES CARACTÉRISTIQUES

Les spécificités précédentes concernant les composantes vont permettre un étiquetage très rapide. Un masque d'analyse tel que celui présenté figure 7, associé à un balayage indiqué sur l'exemple de la figure 8, permet d'étiqueter chaque pixel de manière purement causale (uniquement en fonction des pixels déjà traités). Cet étiquetage, tirant profit de la structure particulière des composantes, est donc à la différence des méthodes traditionnelles, très simple et facilement implantable sur une architecture matérielle.

De plus, ce procédé permet de coder les composantes en même temps qu'elles sont étiquetées, à l'aide d'un codage de Freeman

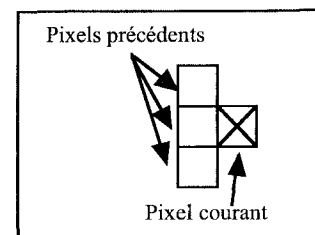


Figure 7. – Masque d'analyse.

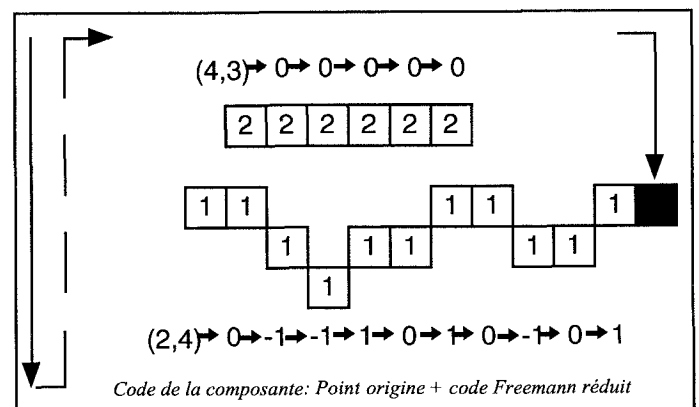


Figure 8. – Procédé d'étiquetage dans la pyramide horizontale.

réduit. Ainsi une composante connexe sera complètement décrite par son point origine et par la chaîne de codes donnant les pentes locales de la composante.

En un seul passage, l'ensemble des composantes est donc étiqueté, codé, et un certain nombre de caractéristiques telles que le centre de gravité, la droite de régression, l'erreur quadratique moyenne par rapport à cette droite, sont également déterminées.

3.4. DISCRIMINATION DES OBJETS

Les composantes connexes sont classées dans une des quatre classes suivantes : *objet court*, *objet irrégulier*, *objet trait*, *objet texte*. La description analytique fournie par la fonction précédente est suffisante pour écarter un certain nombre d'objets de type clairement différent de ceux attendus (*trait* ou *texte*). En particulier, un objet sera classé court si sa longueur est inférieure à L_{min} (typiquement, L_{min} est égale à trois). Un objet est considéré irrégulier lorsqu'il présente une erreur quadratique moyenne importante.

Parmi les composantes restantes, pour distinguer les composantes *trait/texte*, on a recours à la deuxième caractéristique du texte constituant le second attribut : présence de composantes orthogonales. Cette information étant disponible à plus haute résolution que la première, la détection des transitions dans le voisinage d'une ligne détectée dans une image d'étiquettes s'effectue sur l'image orthogonale des étiquettes au niveau immédiatement inférieur de la pyramide (image complémentaire).

Le procédé, illustré sur la figure 9, est décomposé en deux phases :

1-**Projection**, où l'image courante est agrandie pour que sa taille soit identique à celle de l'image au niveau inférieur.

2-**Mise en correspondance**, où l'on vient détecter dans l'image complémentaire les transitions perpendiculaires éventuelles.

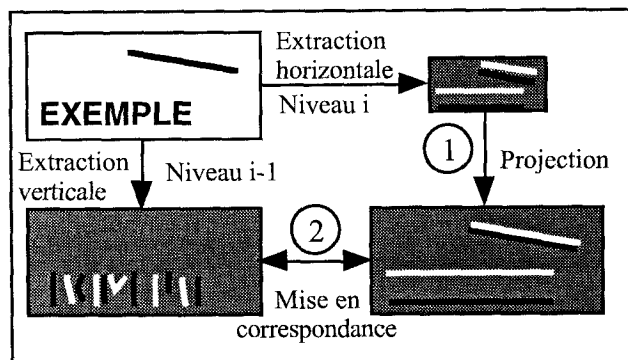


Figure 9. – Schéma de discrimination des objets de type trait/texte.

Si un pixel de l'image courante appartient à un objet de classe *Trait/Texte*, on teste dans son voisinage de l'image complémentaire la présence de pixels objet (supposés alors appartenir au corps d'un caractère). A la fin du balayage de l'image, le rapport entre le nombre de transitions détectées et la longueur de l'objet détermine la classe de cet objet.

4. Résultats de la localisation

Cette méthode originale d'extraction de textes et de traits fut en premier lieu dédiée à la localisation du bloc adresse sur les

paquets postaux. Cette application présentant une large diversité de documents en terme de style de texte et de fonds d'image, les très bons résultats obtenus sur près de trois cents images ont permis de valider l'approche.

Un des avantages apportés par cette méthode est tout d'abord de pouvoir extraire dans le même temps deux types d'objets prépondérants : les textes et les traits. Concernant le texte, aucune limitation n'est imposée ni sur la taille, ni sur la nature. Les lignes inclinées peuvent également être extraites sans traitement supplémentaire. Enfin la méthode se révèle rapide. La première raison en est la relative faible résolution des images de départ (75 dpi au lieu des 200 dpi communément utilisés). La seconde raison est l'utilisation de traitements locaux et réguliers.

L'implémentation des algorithmes a été réalisée sur une machine Sparc 2. Concernant les deux étapes les plus complexes, l'étape de Ternarisation-Squelettisation est effectuée en un temps moyen de 420 ms pour une image 512*512, alors que la phase de Labelisation-Codage est achevée en 196 ms, pour un nombre moyen de 1500 objets.

Les traitements dits de haut niveau permettent tout d'abord de retrouver l'intégralité du texte par combinaison des bords supérieurs et inférieurs, d'extraire ensuite les blocs de lignes structurés. Les traitements ayant été réalisés de manière indépendante sur les niveaux de la pyramide, une phase finale de résolution de conflits classe tous les blocs extraits pouvant correspondre à un BAD pour l'étape suivante de segmentation. La généricité du système d'analyse a pu être vérifiée en l'utilisant pour d'autres applications telles que la localisation de codes à barre et l'interprétation logique de documents (extraction des paragraphes, lignes isolées, titres, traits)[12].

Les figures 10 à 12 fournissent des exemples de localisation de lignes de texte. Elles montrent l'extraction des bords inférieurs (a), des bords supérieurs (b) et de la ligne médiane résultante (c).

5. Binarisation par zone d'intérêt

Classiquement en traitement d'images de documents, la segmentation est effectuée après la binarisation. L'approche présentée est différente : nous utilisons la phase de localisation des zones d'intérêt sur l'image multi-niveaux de gris comme une macro-segmentation. Cela permet ensuite de considérer chaque zone d'intérêt comme une sous-image en niveaux de gris. Le seuil de binarisation est ainsi défini très finement sur les seules informations pertinentes.

En effet, sur les objets postaux, mais ce n'est pas le seul cas, il est courant de rencontrer des images possédant plusieurs gammes de variations des niveaux de gris, ou encore des images présentant un fort bruit dû aux reflets de plastiques transparents. La méthode proposée s'affranchit de ces inconvénients ; le nombre de pixels objet n'est plus négligeable devant le nombre de pixels fond, ces deux classes sont alors plus facile à séparer.

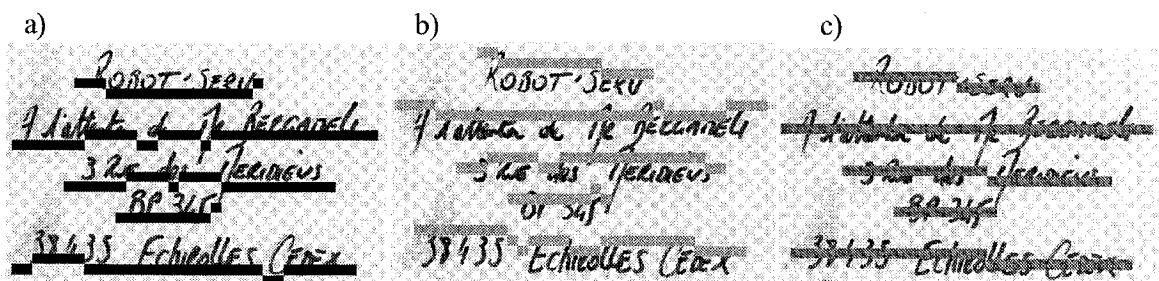


Figure 10. – Texte manuscrit, extraction niveau 3 (9 dpi).



Figure 11. – Texte imprimé incliné, extraction niveau 1 (37,5 dpi).

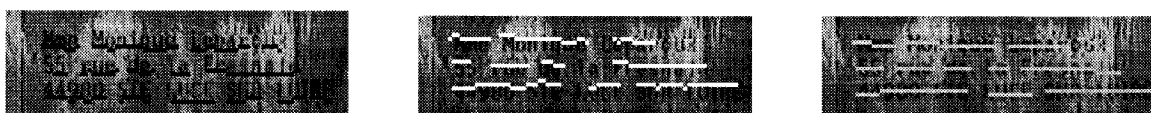


Figure 12. – Texte imprimé bruité, extraction niveau 1 (37,5 dpi).

Pour calculer le seuil de binarisation, il existe diverses méthodes. Nous avons choisi la méthode développée dans [13] qui permet de calculer automatiquement un seuil à partir de l'histogramme des niveaux de gris; le seuil est la valeur qui maximalise la variance inter-classe. Cette méthode permet de bien mettre en évidence les apports du principe de la coopération entre segmentation et binarisation.

L'exemple suivant illustre bien le concept. La figure 13 présente une image en multi-niveaux de gris typique du domaine d'application. Un agrandissement du Bloc Adresse Destinataire (BAD) est effectué figure 14. Cette zone d'intérêt a pu être sélectionnée à partir des informations fournies par la méthode de localisation décrite précédemment, l'étape de formation et de sélection du Bloc Adresse étant décrite dans [14]. Une observation rapide de ces deux images relève des valeurs de gris très disparates, introduites notamment par des reflets sur le film plastique de couverture. Nous avons appliqué la méthode à cette image en considérant trois types de zones d'intérêt : l'image complète, le BAD, la ligne. La figure 15 présente le résultat classique de la binarisation à partir d'un seuil calculé sur l'histogramme de tous les pixels de l'image, la figure 16 sur l'histogramme du BAD, la figure 17 sur l'histogramme de la ligne « rue », la figure 18 sur l'histogramme de la ligne « ville ».

Les différences entre une binarisation globale et le principe présenté sont assez évidentes sur ces images. Nous constatons visuellement une importante amélioration entre la figure 15 et la figure 16. La binarisation sur la zone ligne par rapport à la zone

BAD permet la segmentation fine des caractères qui ne sont plus liés et qui présentent des boucles creusées.

La démonstration de l'intérêt d'avoir localisé les zones sur l'image multi-niveaux aurait pu se faire avec une autre méthode que celle d'OTSU. On peut par exemple utiliser la technique proposée par [15], où le seuil est déterminé à partir de l'histogramme calculé à partir des niveaux de gris des pixels induisant des extrema sur les images des dérivées secondes à calculer préalablement.

6. Traitement de post-segmentation

6.1. STRUCTURE GÉNÉRALE

Cette dernière étape va permettre de corriger les imperfections inhérentes soit à la méthode même de localisation soit à des configurations locales ambiguës. La méthode de localisation proposée précédemment mettait en évidence avant tout un caractère macroscopique linéique pour localiser les lignes de texte, ainsi les extensions de caractères ne sont pas la plupart du temps dans la zone localisée, inversement et quelle que soit la méthode de localisation mise en œuvre, des éléments parasites situés dans la zone localisée peuvent être présents. La post-segmentation doit être capable de gérer ces deux types de situations : perte d'information et prise en compte de composantes perturbantes dans la

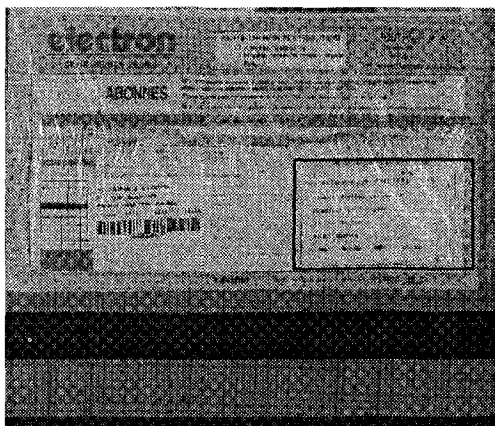


Figure 13. – Image test en niveaux de gris.

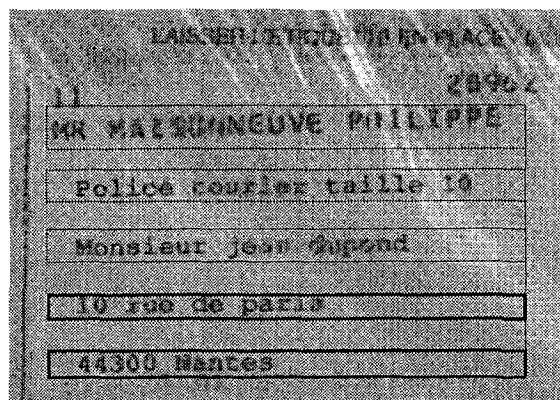


Figure 14. – Agrandissement de la zone du BAD.

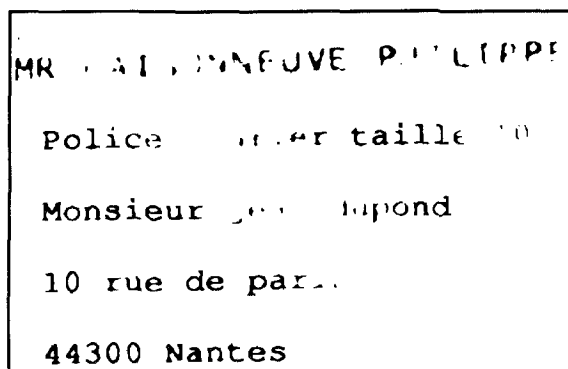


Figure 15. – Binarisation sur toute l'image. Seuil calculé : 101.

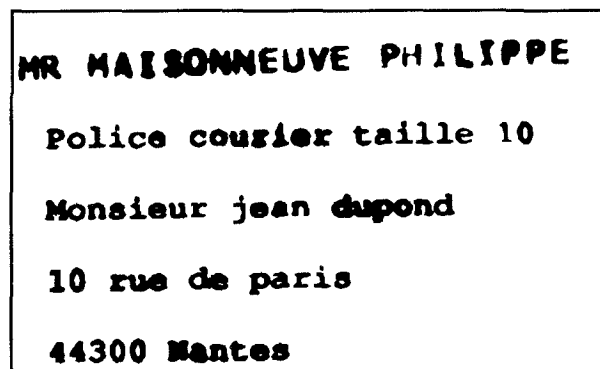


Figure 16. – Binarisation sur le BAD. Seuil calculé : 146.

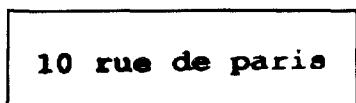


Figure 17. – Binarisation sur la ligne "rue". Seuil calculé : 140.

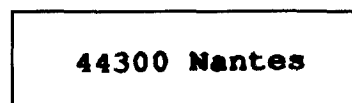


Figure 18. – Binarisation sur la ligne "ville". Seuil calculé : 133.

zone localisée. Ces traitements sont évidemment vitaux pour une bonne reconnaissance, en particulier sur du texte manuscrit pour séparer les lignes se chevauchant par des extensions de caractères.

A titre d'exemple, le traitement de post-segmentation doit effectuer les opérations suivantes sur les lignes de la figure 19 qui représente le résultat à l'issue de la localisation :

- ligne 4, la partie supérieure du *R* débordant du rectangle doit être réintégrée à la ligne afin d'en assurer la bonne reconnaissance,
- le *B* de Beccameli de la ligne 3 et le *M* de Méridiens de la ligne 2 doivent être séparés,
- ligne 1, l'extension du *M* doit être éliminée pour que le système de lecture reconnaisse BP345 au lieu BP de 3451.

La post-segmentation traite chaque ligne indépendamment des autres. La procédure qui l'effectue, dont la structure est présentée figure 20, comprend plusieurs fonctions qui réalisent les opérations suivantes :

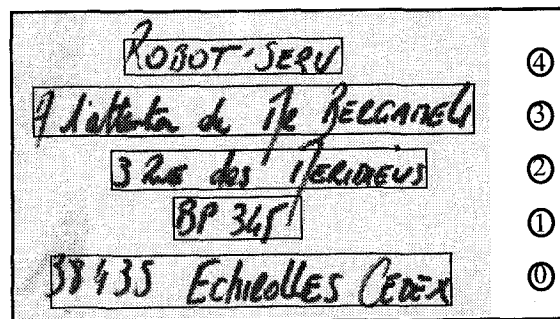


Figure 19. – Résultat de la localisation.

- extraction des composantes débordant à l'extérieur du rectangle de la ligne par suivi de contours, suivant le seuil de binarisation,
- classification de ces composantes,
- étiquetage des composantes connexes de la ligne,

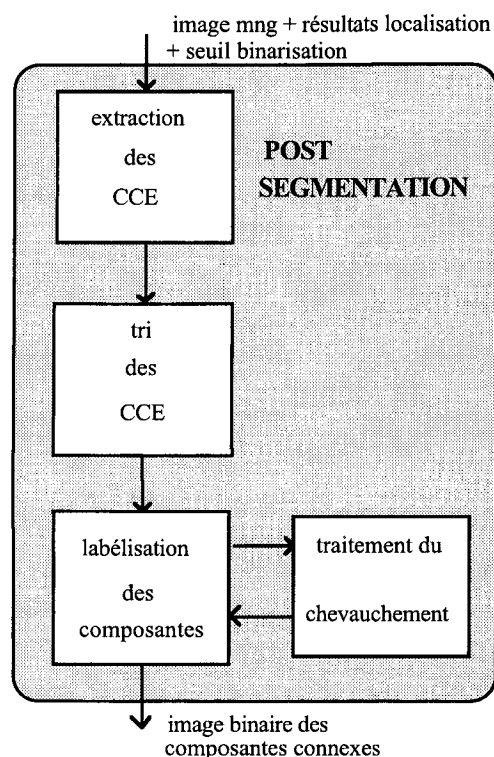


Figure 20. – Structure de la procédure de post-segmentation des lignes.

- fractionnement des composantes issues d'un chevauchement entre deux lignes.

Avant de décrire les fonctions de ces différents blocs, il est nécessaire de préciser au préalable quelques définitions et notations utilisées pour décrire la suite des traitements.

RL : Rectangle Ligne; il est défini par le module de localisation des zones d'intérêt. Représenté par les coordonnées de deux points extrêmes, il contient le corps d'une ligne de l'adresse.

CCE : Composante Connexe d'Échappement, sa définition est la suivante :

- soit R_i la région de pixels définie par le RL de la ligne étudiée i ,
- soit Sb_i le seuil de binarisation calculé sur la région R_i ,
- alors la $n^{ième}$ CCE C_i^n , attachée à la ligne i est composée des pixels connexes p_k , tels que :

$$\forall p_k \in C_i^n, \text{niveau de gris de } p_k < Sb_i,$$
 avec $C_i^n \cap R_i \neq \emptyset$ et $C_i^n \not\subset R_i$.

De plus, $\forall m \neq n$, on a $C_i^m \cap C_i^n = \emptyset$ (les composantes sont de taille maximale). Nous distinguons deux types de CCE, les CCE simples et les CCE complexes :

Soient R_j les régions délimitées par les rectangles RL des lignes autres que la ligne étudiée i , alors la CCE C_i^n est :

- simple si $\forall j \neq i, C_i^n \cap R_j = \emptyset$
- complexe si $\exists j \neq i, C_i^n \cap R_j \neq \emptyset$.

Dans ce cas, les lignes j telles que $C_i^n \cap R_j \neq \emptyset$ sont appelées les lignes d'échappement de la ligne i .

Il est important de souligner qu'une CCE est totalement liée à une ligne et à son seuil de binarisation. En conséquence, un objet ayant visuellement dans l'image en niveaux de gris des pixels dans deux RL, pourra générer deux CCE différentes selon les valeurs du seuil de binarisation propre à chacune des deux lignes.

Une CCE complexe peut être :

- avec chevauchement si elle résulte du chevauchement de caractères appartenant à des lignes différentes,
- sans chevauchement si elle résulte seulement d'un prolongement de caractère hors du RL de la ligne étudiée dans le RL d'une autre ligne ou l'inverse.

Afin d'illustrer ces différentes définitions, nous pouvons reprendre l'exemple de la figure 19. Soit la ligne $i = 2$ considérée comme la ligne étudiée : elle comporte trois CCE

- le point sur le I de Méridiens est une CCE simple,
- le M de Me est une CCE complexe sans chevauchement,
- le M de Méridiens est une CCE complexe avec chevauchement.

6.2. EXTRACTION ET CLASSIFICATION DES COMPOSANTES CONNEXES D'ÉCHAPPEMENT

L'objectif de ce bloc est d'extraire les CCE et de les classer parmi l'une des trois catégories suivantes : $C1 = CCE$ appartenant totalement à la ligne; $C2 = CCE$ à exclure totalement de la ligne courante; $C3 = CCE$ appartenant en partie à la ligne courante (cas de chevauchement).

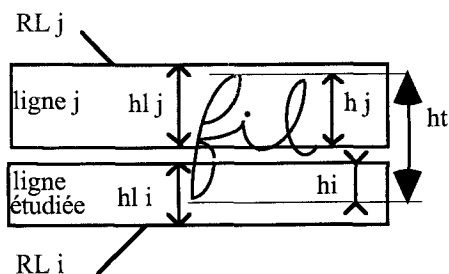
Pour extraire les CCE, le principe est simple, il consiste à parcourir le pourtour du RL et suivre tout objet qui s'en « échappe ». Ce suivi s'effectue dans l'image en niveaux de gris, en utilisant le seuil de binarisation calculé pour la région définie par le RL de la ligne étudiée.

Lors du suivi de contour, nous collectons des données sur la CCE suivie :

- intersection ou non avec un autre RL classant la CCE comme simple ou complexe,
- rectangle dans lequel est inscrite la CCE,
- rectangle encadrant la CCE au sein de la ligne étudiée,
- rectangle encadrant la CCE dans le RL des lignes d'échappement,

Si la CCE obtenue est simple alors elle est automatiquement classée en C1, c'est-à-dire complètement réintégrée à la ligne de laquelle elle s'échappe. Si la CCE est complexe, un arbre de décision faisant intervenir des paramètres géométriques et topologiques procède au classement suivant l'une des trois classes.

Les critères de classement font intervenir quatre rapports calculés à partir des grandeurs présentées figure 21 qui sont :



coef_haut= hi/ht
 coef_prof= hi/hli
 rap_haut(j)= h_j/hi
 rap_prof(j)= $(h_j/hlj)/(hi/hli)$

Figure 21. – Grandeurs utilisées pour le classement.

Deux bases d'images ont été utilisées, l'une de 62 images pour régler expérimentalement les seuils de décisions, l'autre de 171 images pour évaluer la robustesse de la méthode de classification, table 1. Le résultat du classement des CCE complexes dans l'une des trois classes : appartenance à la ligne (C1), non appartenance à la ligne (C2) et chevauchement (C3) se fait sans aucune erreur sur chacune de ces deux bases.

Tableau 1. – Classification des composantes connexes d'échappement

	Base Apprentissage	Base Test
Nombre d'images	62	171
Nombre de lignes traitées	194	673
Nombre de CCE complexes	C1 : 12 C2 : 72 C3 : 4	C1 : 64 C2 : 103 C3 : 16

Les composantes 8-connexes internes au Rectangle Ligne et les CCE sont alors extraites suivant le seuil de binarisation calculé précédemment et étiquetées. Si une CCE possède l'attribut « avec chevauchement » (C3), alors le bloc de traitement du chevauchement est activé.

Les résultats de ce dernier traitement permettent de créer l'image binaire finale par ligne.

6.3. TRAITEMENT DU CHEVAUCHEMENT

Nous avons déjà défini un chevauchement comme le résultat du croisement des traits de caractères de lignes différentes. S'appuyant sur cette définition, il est alors intéressant d'extraire de l'image binaire de la CCE une représentation symbolique faisant apparaître sur un graphe des arcs représentatifs des traits et des sommets représentatifs des croisements.

La première partie de ce traitement consiste donc à construire un pseudo-squelette. A partir du graphe descriptif de ce pseudo-squelette, une identification de la configuration du chevauchement

est tentée. Si la configuration correspond à un cas du dictionnaire des chevauchements alors la fragmentation est effectuée de manière adaptée.

6.3.1. Pseudo-squelettisation

Les différents concepts utilisés reposent sur les travaux de [16] et [17]. Il s'agit d'extraire de l'image binaire les squelettes des traits. Tout ce qui n'est pas identifié comme trait constitue des régions complémentaires.

Un trait est une forme dont la largeur reste régulière et est inférieure à la longueur. Ces deux critères vont piloter l'extraction des traits. Schématiquement, le principe consiste à balayer l'image pour chercher des germes de trait. A partir d'un pixel contour, la séquence de pixels la plus longue à l'intérieur du trait jusqu'au prochain pixel contour indique la direction principale du trait. Une coupe germe est alors choisie perpendiculairement à la direction principale, supposée de ce fait être sur la largeur du trait. Cette coupe est un segment de droite joignant un pixel du contour droit et un pixel du contour gauche.

A partir de là, le but est de progresser le long des contours droit et gauche, en choisissant les coupes de longueur minimale, et en vérifiant le respect des critères de progression.

Un point squelette est placé dans R2 pour chaque coupe retenue, sur l'axe médian des contours. L'extraction du squelette du trait s'interrompt lorsque la largeur du trait évolue hors d'un gabarit défini à l'initialisation de la coupe germe.

Ainsi, pour chaque squelette de trait, nous disposons d'une liste de coupes transversales au trait, caractérisées chacune par deux pixels contour (droit et gauche) et un point squelette.

Après l'extraction des traits, la détection des régions complémentaires est effectuée par simple complémentarité. Les pixels contours libres (n'appartenant pas à un trait) sont suivis à partir de chaque extrémité de traits permettant ainsi l'extraction et la caractérisation des régions complémentaires.

Lors de ces deux opérations, extraction de squelettes de traits puis extraction de régions complémentaires, différentes caractéristiques sont collectées :

- rectangles englobant les traits, coordonnées des points squelettes et des points contours associés, labels des régions complémentaires de chaque extrémité de trait,
- rectangles englobant les régions complémentaires, poids de chaque région complémentaire (nombre de traits y aboutissant), labels des traits aboutissant.

La figure 22-a présente une CCE extraite de l'exemple de la figure 19 ligne 2 et son pseudo-squelette en 22-b. Son graphe est présenté figure 23.

6.3.2. Fragmentation

Bien que le nombre de chevauchements différents soit quasiment infini, les cas génériques représentatifs d'un grand nombre de

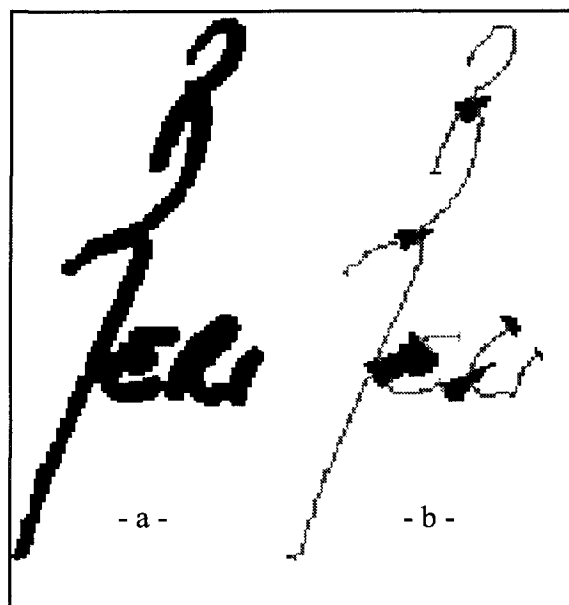


Figure 22. – CCE et son pseudo-squelette.

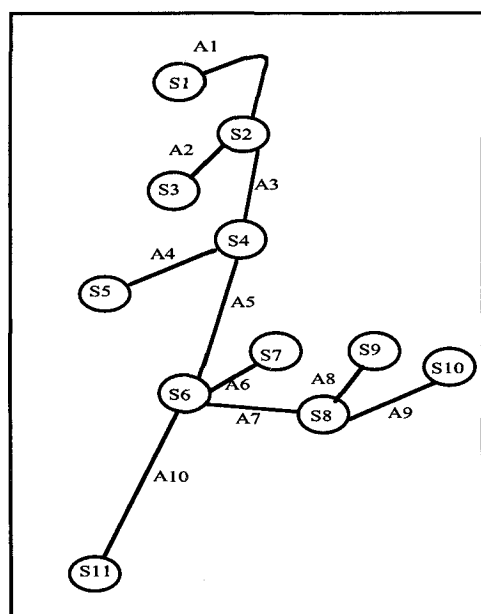


Figure 23. – Graphe structurel associé à la figure 22.

configurations ont été décrits dans un dictionnaire des chevauchements. Celui-ci contient 128 configurations obtenues par combinaison de deux types d'extensions : segment de droite et boucle suivant la localisation du chevauchement. Quelques uns de ces cas, les plus représentatifs, sont illustrés figure 24.

A partir du graphe structurel de la composante, l'identification du chevauchement s'effectue en considérant : le nombre d'arcs d'échappement, la localisation des sommets, le degré de ces sommets. Lorsque la classe est identifiée, nous savons si l'extension sortant du Rectangle Ligne est une boucle ou un segment de droite.

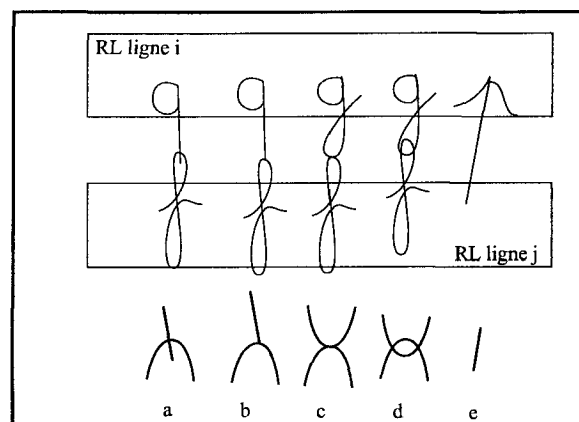


Figure 24. – Exemples du dictionnaire des chevauchements.

Les traitements suivants consistent à rechercher les éléments constitutifs du segment ou de la boucle et à éliminer le reste. Ainsi, en particulier une étude du prolongement inter-arc peut être demandée. Le principe de cette fonction consiste à calculer les tangentes des squelettes de traits aboutissant au croisement, pour cela une approximation polynomiale des squelettes de traits est effectuée, la comparaison des pentes permet de déterminer si un arc est bien dans le prolongement d'un autre. A chaque prise de décision d'effacement, les pixels correspondant sont mis à zéro dans l'image binaire de la CCE.

Les cas traités sont des chevauchements qui engendrent une intersection (24-a, 24-b, 24-c) ou deux intersections (24-d) de traits inter-caractères. Le cas 24-e est particulier dans le sens où il n'y a pas de chevauchement, il permet de traiter une éventuelle erreur de classification. Les cas plus complexes nous semblent plus critiques à déceler et à fragmenter. Ils sont de toute manière peu fréquents.

Dans l'exemple du graphe figure 23, S4 est reconnu comme sommet de chevauchement (cas 24-b), A4 et A5 comme des arcs d'échappement, le graphe est fragmenté au niveau du nœud S4, éliminant de la ligne 2 les arcs A1 à A3.

La figure 25 présente le résultat final de la post-segmentation sur la ligne 2 de la figure 19. Ce résultat est à comparer à celui

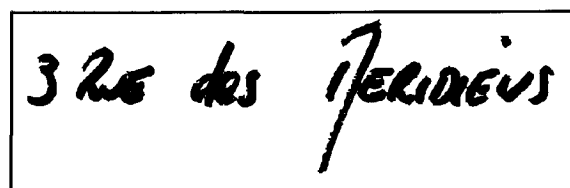


Figure 25. – Image binaire finale.

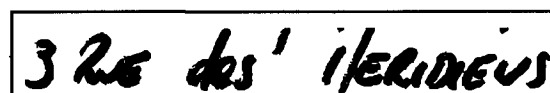


Figure 26. – Image binaire sans post-segmentation.

de la figure 26 qui présente ce que serait la même ligne sans la post-segmentation.

Le nombre de cas de chevauchement dans les images de la base étant faible (20 CCE), toutes les situations n'ont pu être testées. Sur les cas rencontrés, 18 sont traités de manière correcte par cette méthode; correcte signifie ici que l'image fragmentée restitue visuellement l'idée que nous nous faisons du caractère sans le chevauchement. Deux cas sont mis en défaut, ces deux erreurs proviennent en fait d'une mauvaise localisation du chevauchement : le sommet présent dans l'interligne n'est pas le sommet générateur du chevauchement.

7. Conclusions

Nous avons présenté une méthode permettant, à partir d'une image complexe en niveaux de gris, de fournir une image binaire par ligne, directement exploitable par un système de reconnaissance. Bien que le couplage avec un système de lecture ne soit pas encore réalisé pour évaluer quantitativement la qualité de l'extraction des lignes, nous pouvons, à partir d'une analyse visuelle, qualifier les résultats obtenus de très bons. L'originalité de l'approche tient d'une part au schéma coopératif localisation-extraction, d'autre part, elle est basée sur les propriétés issues de la perception de l'écriture où celle-ci est considérée non pas de façon « microscopique » mais « macroscopique » en basse résolution ce qui permet une très bonne localisation du texte. Lignes de texte et traits se trouvant être des éléments essentiels, et souvent mêlés dans des documents, une démarche unique pour leur extraction se révèle fort judicieuse. Elle permet de s'affranchir d'une extraction systématique de toutes les composantes connexes individuelles à la différence de nombreuses méthodes actuelles. On peut également indiquer que, bien que deux directions, horizontale et verticale, aient simplement été utilisées, les filtres ont une sélectivité angulaire suffisamment réduite pour pouvoir détecter les objets présentant des inclinaisons importantes (jusqu'à $\pm 45^\circ$). Cet ensemble de traitements bas-niveaux fournit des descriptions analytiques et logiques des objets. Les phases suivantes sont bien sûr directement liées à l'application : ainsi, sans que cela soit développé dans cet article, il a été possible de s'en servir pour la localisation de code-barres ou encore du bloc-adresse destination sur des objets postaux classés difficiles. Il est à noter enfin que grâce à la régularité des traitements, tous les algorithmes de

bas niveaux peuvent être projetés sur des structures matérielles de type parallèle, la référence [12,18] décrivant les architectures matérielles supportant ces traitements.

BIBLIOGRAPHIE

- [1] A. Chéhikian, « Binarisation d'images, deux solutions à ce problème », *Traitement du Signal*, Vol. 6, n°1, 1989, p. 59-68.
- [2] Y. Tang, C.Y. Suen, « Document analysis and understanding : a brief survey », *Proc. First ICDAR*, Saint Malo, France, 1991, pp. 17-31.
- [3] G. Nagy, S. Seth, « Hierarchical representation of optically scanned documents », *Proc. Seventh ICPR*, Montreal, Canada, 1984, pp. 347-349.
- [4] F. Hönes, J. Litcher, « Layout extraction of mixed mode documents », *Machine Vision and Applications*, Vol. 5, 1994, pp. 237-246.
- [5] L. Likforman-Sulem A. Hanimyan, C. Faure, « A Hough based algorithm for extracting text lines in handwritten documents », *Proc. Third ICDAR*, Montreal, Canada, Aug. 1995, pp. 774-777.
- [6] J. Labiche, J.L. Bourdon, Y. Lecourtier, « A cognitive strategy to achieve an automatic extraction of character areas in noisy images », *Transputer '92*, Arc-et-Senans, France, May 1992, pp. 103-113.
- [7] P.J. Burt, E.H. Adelson, « The Laplacian pyramid as a compact Image Code », *IEEE Trans. on Communications*, Vol. COM-31, 1983, pp. 532-540.
- [8] C. Viard-Gaudin, D. Barba, « Robust extraction of destination address block using a multi-level segmentation », *Proc. Ninth ATC USPS*, Washington, USA, Dec.1992, pp. 1257-1270.
- [9] D.G. Lowe, « Organisation of smooth image curves at multiple scales », *IJCV*, Vol. 3, 1989, pp. 119-130.
- [10] A.K. Jain, K. Bhattacharjee, « Text segmentation using Gabor filters for automatic document processing », *Machine Vision and Applications*, Vol. 5, 1992, pp. 169-184.
- [11] P.G. Mulgaonar, A. Bergman, « Address block location : the SRI approach », *Proc. Fifth ATC USPS*, Washington, USA, 1986, pp. 161-178.
- [12] O. Déforges, « Segmentation robuste d'images de documents par une approche multirésolution. Conception et validation d'une architecture parallèle dédiée », *Thèse de doctorat de l'Université de Nantes*, Sept. 1995.
- [13] N. Otsu, « A threshold selection method from gray-level histogram », *IEEE Trans. on Syst. Man and Cyb.*, Vol. SMC-9, n°1, 1979, pp. 62-66.
- [14] O. Déforges, D. Barba, « A robust and multiscale document image segmentation for block line/text line structures extraction », *Proc. Twelfth ICPR*, Jérusalem, Israël, Oct. 1994, pp. 306-310.
- [15] T. Pavlidis, « Threshold selection using second derivatives of the gray scale image », *Proc. Second ICDAR*, Tsukuba, Japan, Oct.1993, pp. 274-276.
- [16] J.C. Pettier, « Extraction d'une représentation adaptée à la reconnaissance de l'écriture », *Thèse de l'Université de Rennes I*, 1994.
- [17] T. Paquet, T. Vallee, Y. Lecourtier, « Extraction de primitives par suivi de traits dans l'image binarisée d'un mot manuscrit », *Bigre n°68, Colloque RAE*, Le Havre, 1990, p. 179-188.
- [18] G. Ramstein, O. Déforges, P. Bakowski, « A design tool for the specification and the simulation of array processor architectures », *IEEE Inter. Conf. on Arch. Spec. Array Processors (ASAP)*, Strasbourg, France, Jul. 1995, pp. 322-329.

Manuscrit reçu le 30 Janvier 1995.

LES AUTEURS

Olivier DÉFORGES

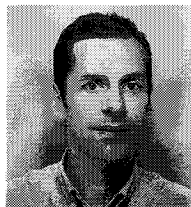


Olivier Déforges est né en 1969. Ingénieur de l'Institut de Recherche et d'Enseignement Supérieur aux Techniques de l'Électronique IRESTE (1992), il est également titulaire d'un Doctorat de l'Université de Nantes et de l'École Centrale de Nantes, en Sciences pour l'Ingénieur dans la spécialité Automatique et Informatique Appliquées (1995). Ses travaux concernent à la fois l'analyse d'images de documents et la conception d'architectures matérielles dédiées.

Patrice PIQUIN

Patrice Piquin a obtenu un diplôme d'ingénieur (1994) du Conservatoire National des Arts et Métiers dans la spécialité Automatismes Industriels. Son travail de mémoire ingénieur (12 mois) s'est déroulé au sein du laboratoire SEI de l'IRESTE où il a travaillé sur la segmentation fine et la classification de blocs de lignes en vue de leur lecture. Il est maintenant responsable de recherches et développements chez Zircotube.

Christian VIARD-GAUDIN



Christian Viard-Gaudin est Maître de Conférences au département Génie Électrique et Informatique Industrielle de l'IUT de Nantes. Il a obtenu un diplôme d'ingénieur de l'ENSM (devenue École Centrale de Nantes) en automatique puis une thèse de troisième cycle en 1981 sur la commande adaptative. Il a ensuite contribué à la mise en place d'un nouveau laboratoire de recherche en Analyse et Traitement des Images (1986) à la naissance de l'IRESTE, celui-ci

est maintenant intégré à l'équipe Systèmes Électroniques et Informatiques (SEI). Ses activités de recherche sont centrées sur l'analyse d'images et plus particulièrement sur le traitement d'images de documents.

Dominique BARBA

Dominique Barba est né en France en 1944. Il est titulaire d'un doctorat de 3ème cycle en télécommunications de l'Université de Rennes I obtenu en 1969 et d'un doctorat es Sciences Mathématiques (mention informatique) de l'Université de Paris VI sur un sujet lié au traitement numérique des images avec critère psychovisuel de qualité. Il est actuellement professeur à l'IRESTE (Université de Nantes) et responsable d'une équipe de recherche en Analyse et Traitement des images au sein du laboratoire Systèmes Électroniques et Informatiques (SEI). Ses centres d'intérêt concernent le traitement numérique des images fixes et animées, spécialement en vue de l'analyse ou du codage des images.